Technical Research Note 236

AD

A COMPARISON OF TWO METHODS OF GISTING

Joyce L. House

Combat Systems Research Division

BESRL

U. S. Army

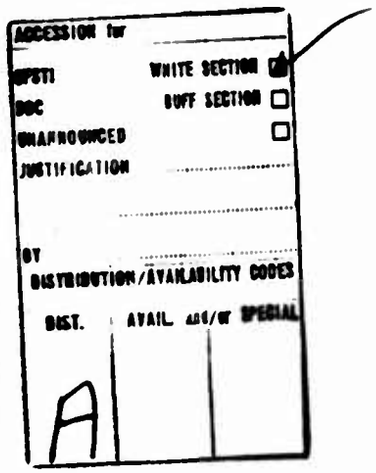
Behavior and Systems Research Laboratory

March 1972

BEHAVIOR AND SYSTEMS RESEARCH LABORATORY

An activity of the Chief, Research and Development

J. E. UHLANER
Director



NOTICES

DISTRIBUTION: Primary distribution of this report has been made by BESRL. Please address correspondence concerning distribution of reports to: U. S. Army Behavior and Systems Research Laboratory, Attn: RDMR-BLZ, 1300 Wilson Boulevard, Arlington, Virginia 22209.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the Behavior and Systems Research Laboratory.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

DD Form 1473

13. Abstract - continued

the amount of time spent on each presentation was controlled; 3) in both methods, repetition produced gains in number of key items correctly reported and in overall quality of the gist. In summary, the results imply that on-line gisting is a feasible and economical consideration for timely message contents reporting; the best method of gisting varies with type of material; and that there is a need to identify the properties of a message which determines its difficulty in order to choose the most effective gisting method for use in a given communication.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Behavior and Systems Research Laboratory, Arlington, Va.	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

A COMPARISON OF TWO METHODS OF GISTING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

Joyce L. House

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
November 1971	31	4

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. DA R&D PJ No. 2Q024701A723 c. Monitor Skills d. d-03	Technical Research Note 236 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	US Army Security Agency, Arlington, Va.

13. ABSTRACT

The primary objective of research conducted by the MONITOR SKILLS Work Unit of the Behavior and Systems Research Laboratory is to improve performance in Army monitor jobs with special emphasis on developing and testing new work methods. An important segment of the research is concerned with human factors studies in communication analysis processing. The present technical research note reports on two experimental methods of gisting -- reporting in brief form the essential information in a communication -- in comparison with baseline performance in simulated on-line gisting. Twenty-four communications processors were each tested using two methods of gisting. In one method (free repeat), the processor listened once to the complete message with no option to stop or replay, preparing as complete a gist as possible, and was then free to replay the tape as he judged necessary to complete the gist. In the second method (forced repeat), the processor listened to the complete message five additional times with no option to stop or replay. The resulting gists were evaluated in terms of key items of information correctly reported, quality of gist as rated by four judges, and in the case of the free repeat method the number of times the tape was stopped and replayed in the gisting process.

The two work methods of gisting examined in this study were designed so that the effect of repetition on the accuracy and quality of the gist could be determined. Results of the experiment showed 1) the free repeat method produced a higher quality gist and a higher percentage of key items correctly reported; 2) less time was used in producing a complete gist in the free repeat method; however, in the forced repeat method,

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
*Communications research						
*Monitor performance						
*Gisting						
voice processing procedures						
signal detection apparatus						
simulated on-line gisting						
*Repetition methodology						
Real-time gisting						
Transcription						
*Gisting work methods						
Military communications systems						
Communication processors						
Forced repeat .						
Free repeat						

Technical Research Note 236

AD

A COMPARISON OF TWO METHODS OF GISTING

Joyce L. House

Michael Kaplan, Senior Task Leader

COMBAT SYSTEMS RESEARCH DIVISION
Aaron Hyman, Chief

BEHAVIOR AND SYSTEMS RESEARCH LABORATORY

Office, Chief of Research and Development
Department of the Army

1300 Wilson Boulevard, Arlington, Virginia 22209

March 1972

Army Project Number
2Q024701A723

Monitor Skills d-03

BESRL Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The complex weapons and surveillance systems of the modern Army have created a relatively new series of jobs requiring operators to monitor instrument panels, radarscopes, communication nets, and other types of signal detection apparatus. The MONITOR SKILLS Work Unit of the Behavior and Systems Research Laboratory deals with the many personal, environmental, and situational variables affecting human performance in the detection and analysis of a broad variety of signals. The primary objective is to improve performance in Army monitor jobs, with special emphasis on developing and testing new work methods.

An important segment of the research is devoted to human factors studies in communication analysis and processing. Technical Research Note 236 reports on two experimental methods of gisting--that is, abstracting the essential elements of a voice message--in comparison with baseline performance in simulated on-line gisting.

The entire work unit is responsive to objectives of RDT&E Project 2Q024701A723, "Human Performance Experimentation," FY 1972 Work Program.

J. E. UHLANER, Director
Behavior and Systems
Research Laboratory

A COMPARISON OF TWO METHODS OF GISTING

BRIEF

Requirement:

To evaluate two methods of gisting--reporting in brief form the essential information in a communication--in terms of accuracy in reporting key items and quality of the summary produced.

Procedure:

Twenty-four communications processors were each tested using two methods of gisting. In both methods, the subject listened to the complete message one time with no option to stop or replay, preparing as complete a gist as he could. In one method, he was then free to replay the message, starting and stopping as he judged necessary to complete the gist (free repeat). In the other method (forced repeat), the subject listened to the complete message five additional times with no option to stop or replay. The resulting gists were evaluated in terms of key items of information correctly reported, quality of gist as rated by four judges, and in the case of the free repeat method the number of times the tape was stopped and replayed in producing a gist.

Findings:

The free repeat method produced a higher quality gist and a higher percentage of key items correctly reported than did the forced repeat method.

The free repeat method also took less time to produce a complete gist. However, using the free repeat method the gister could proceed at his own pace after the first play-through. Using the forced repeat method, the amount of time spent on each presentation was controlled.

In both methods, repetition produced gains in number of key items reported correctly and in overall quality of the gist. Amount of gain varied considerably with the message.

Utilization of Findings:

Complete repetitions of a message with no opportunity to replay segments does not appear to be an economical method of gisting. For some information requirements, and with some messages, on-line gisting or a gist prepared after one or two repetitions may represent an acceptable tradeoff between timeliness and quality of information.

Since messages vary in difficulty, it becomes important to identify the properties which make a message easy or difficult in order to choose the most effective gisting method for use with a given communication.

A COMPARISON OF TWO METHODS OF GISTING

CONTENTS

TABLES

FIGURES

A COMPARISON OF TWO METHODS OF GISTING

A major problem in the field of radio-telephone communications is the rapid extraction of complete and accurate information from the increasingly large volumes of incoming messages. Standard processing techniques are inadequate to cope with this great mass of material in a rapid manner. Currently, communications monitors record incoming messages on magnetic tape and transcribers prepare written verbatim transcripts using these tapes. It has been suggested that a new method be developed to replace this time consuming process. Gisting messages may be a more efficient method of extracting communications information.

Gisting is the process of reporting the content of a communication in abstracted form. On-line or real-time gisting would require the communications monitor to extract the important information from the message as it enters the system. This procedure would result in a saving in personnel requirements as well as in processing time. In addition, the gist would provide in abstracted form all the information needed to be communicated to the next level in the communications chain. Off-line gisting from taped messages would also result in a saving of processing time.

As a standard procedure, gisting places additional requirements on the processor. Gisting demands that the processor, in addition to perceiving and recognizing the words in the message as is required in transcription, filter, organize, and summarize the information these words convey. The best method for gisting from tapes may be quite different from the best method for transcribing. A transcriber frequently replays small sections of a tape in an effort to transcribe every word. In gisting, it is important to listen for thoughts rather than to put down every utterance; it may therefore be more efficient to listen to the entire message or long segments of it.

Two work methods for gisting were examined in the present study. The main purposes were to establish 1) which of two work methods is the better way to gist; 2) the effect of repetition on the accuracy of the gist, and 3) a base-line estimate of the quality and quantity of information extracted from the first presentation of a message.

METHOD

Subjects

Twenty-four communications processors with field experience in transcription of foreign language communications served as subjects.

Test Tape

Two English language communications, referred to as "messages 1 and 2," and each approximately 5 minutes in length, were recorded on separate tapes under studio conditions. The conversations were based on typical foreign language messages. The speakers, although untrained in diction or announcing, made an effort to speak clearly and distinctly. The taped conversations were rated by experienced communications processors as far superior to the average communication.

Apparatus

The tapes were reproduced on Midwestern AN/TNH-11 recorder/reproducers[1] and presented through Telex HTW-2 electrical headsets. The recorder/reproducers were modified so that reversals of the tape could be counted. Counts were recorded on Veeder-Root 120 DC volt electro-mechanical counters. However, some of the modified pieces of equipment introduced a distracting noise and had to be eliminated from the study. Therefore, replay data were not obtained for one-third of the subjects.

Test Procedure

Two gisting work methods were examined in this study--method A, free repeat and method B, forced repeat. In method A, the processor listened to the complete message with no option to stop, start, or replay segments of the message on the first trial. He was requested to record as complete a gist as possible. Following the first trial, the processor had the option to start, stop, and replay the tape as frequently as he wished, until he had finished what he felt was a complete gist of the message.

In method B the processor listened to the complete message six times with no option to stop, start, or replay segments of the communication. With each repetition, the processor was asked to record as complete a gist as possible. Work produced on each trial was available to the processor to add to, correct, or alter during each additional trial.

A repeated measures 2 x 2 x 2 design (method x message x test session) set up as a replicated Graeco-Latin square design was used to assess the effects of gisting methods A and B. Each man was tested using both methods of gisting. The subjects were divided into four

---

[1] Trade names are used only in the interest of precision in reporting. Their use does not constitute indorsement by BESRL or the Army.

groups of six men each. The complete Graeco-Latin square design was:

	Testing Session	
	First	Second
Subjects		
Group I	A1	B2
Group II	B1	A2
Group III	A2	B1
Group IV	B2	A1

A and B stand for methods, 1 and 2 for messages. For example, group I gisted message 1 using method A during the first test session and then gisted message 2 using method B in the second testing session. In this design, each man serves as his own control in comparisons of methods and of messages.

Dependent Variables

Three measures of performance were used:

1. Percent Items Correct. The percent of key items correctly reported in each gist.

2. Quality Rating. Quality was defined in terms of presenting a complete, accurate, and concise summary of all important intelligence information contained in the message. Each gist was rated on a scale from 1 to 5 ranging from unacceptable to superior.

3. Tape Reversals. The number of times the tape was stopped and replayed.

Scoring Method

The gists were evaluated by a panel of four judges selected on the basis of high performance and many years of experience in the field of communications processing. The panel first identified the key items of importance in each conversation. Using this list of items, each judge independently evaluated the work of each man tested. Each judge checked the key item as right, wrong, or omitted, and rated the complete gist on quality. For each man tested, the scores from the four judges were averaged to form a single mean score for each gist for each of the first two performance measures. The reliability of the mean score across judges, ignoring differences in level between judges, was determined by an analysis of variance procedure[2] for each performance score. These reliability

---

[2] Winer, B. J. Statistical principles and experimental design. New York: McGraw-Hill. 1962, p. 131.

coefficients for percent items correctly reported and for quality rating are shown in Table 1.

Table 1

RELIABILITY COEFFICIENTS OF THE MEAN OF THE FOUR JUDGES FOR PERCENT ITEMS CORRECTLY IDENTIFIED AND QUALITY RATINGS OF EACH TRIAL FOR METHODS A AND B.

		Group Reliability	
		Quality Rating	%Items Correctly Reported
Method A			
	Trial 1	.31	.97
	Trial 2	.82	.91
Method B			
	Trial 1	.60	.97
	Trial 2	.60	.96
	Trial 3	.77	.96
	Trial 4	.80	.96
	Trial 5	.79	.95
	Trial 6	.75	.92

RESULTS

Work Method

Comparison of performance for work method A, free repeat, and work method B, forced repeat, revealed that method A was generally superior to method B (Figure 1)[3]. A method x session x message x groups of subjects analysis of variance comparing the first trial of the two methods, and comparing the final repetition of method A with each repetition of method B, was computed separately for each performance measure[2]. Analyses of variance for percent items correctly reported and quality ratings are presented in Tables 2 and 3.

[2] See footnote 2, page 3.

[3] The frequency distribution for performance scores is shown in the Appendix.

Table 2

ANALYSES OF VARIANCE COMPARING THE FIRST TRIAL OF METHOD A WITH THE FIRST TRIAL OF METHOD B AND THE FINAL TRIAL OF METHOD A WITH EACH TRIAL OF METHOD B FOR PERCENT ITEMS CORRECTLY IDENTIFIED

		TRIAL											
		1A vs 1B			Final A vs 1B			Final A vs 2B			Final A vs 3B		
Source of Variance	df	SS	MS	F	SS	MS	F	SS	MS	F	SS	MS	F
Between													
Groups	3	4599.8	1533.3	5.99*	4434.1	1478.0	7.36*	4016.3	1338.8	9.91*	3026.3	1008.7	7.39*
Session x method	1	982.2	982.2	3.84	100.5	100.5		374.1	374.1	2.77	624.3	624.3	4.57*
Session x message	1	76.5	76.5		9.7	9.7		136.7	136.7	1.01	118.2	118.2	.87
Method x message	1	3541.7	3541.7	13.85*	4323.9	4323.9	21.53*	3505.5	3505.5	25.94*	2283.9	2283.9	16.72*
Error	20	5116.8	255.8		4016.31	200.8		2702.4	135.1		2731.7	136.6	
Within													
Session	1	438.9	438.9	2.91	.1	.1		94.1	94.1	.53	235.4	235.4	1.07
Method	1	.4	.4		10,366.6	10,366.6	72.9*	4848.2	4848.2	27.16*	2780.1	2780.1	12.62*
Message	1	1362.4	1362.4	9.04*	940.1	940.1	6.6*	581.1	581.1	3.26	161.0	161.0	.73
S x M x M	1	13.3	13.3		66.9	66.9		.1	.1		.1	.1	
Error	20	3013.0	150.7		2841.3	142.1		3569.9	178.5		4406.6	220.3	

		TRIAL								
		Final A vs 4B			Final A vs 5B			Final A vs 6B		
Source of Variance	df	SS	MS	F	SS	MS	F	SS	MS	F
Between										
Groups	3	2492.2	830.7	6.36*	2029.6	676.5	4.85*	1892.9	631.0	3.20*
Session x method	1	445.4	445.4	3.41	184.9	184.9	1.32	115.4	115.4	.59
Session x message	1	321.6	321.6	2.46	285.3	285.3	2.04	151.3	151.3	.77
Method x message	1	1725.8	1725.8	13.21*	1559.6	1559.6	11.17*	1626.4	1626.4	8.25*
Error	20	2612.2	130.6		2792.0	139.6		3940.7	197.0	
Within										
Session	1	131.6	131.6		15.7	15.7	.11	1.8	1.8	.02
Method	1	1333.7	1333.7	7.11*	530.8	530.8	3.58	209.2	209.2	2.15
Message	1	41.6	41.6		19.3	19.3	.13	25.0	25.0	.26
S x M x M	1	44.5	44.5		31.5	31.5	.21	1.5	1.5	.02
Error	20	3750.9	187.5		2968.7	148.4		2058.6	102.93	

*p < .05

Table 3

ANALYSES OF VARIANCE COMPARING THE FIRST TRIAL OF METHOD A WITH THE FIRST TRIAL OF METHOD B AND THE FINAL TRIAL OF METHOD A WITH EACH TRIAL OF METHOD B FOR QUALITY RATING

		TRIAL											
		[illegible] vs 1B			Final A vs 1B			Final A vs 2B			Final A vs 3B		
Source of Variance	df	SS	MS	F	SS	MS	F	SS	MS	F	SS	MS	F
Between													
Groups	3	3.28	1.09	14.26*	4.21	1.40	4.59*	4.95	1.65	8.14*	3.65	1.22	4.16*
Session x method	1	1.69	1.69	22.05*	.88	.88	2.88	1.17	1.17	5.78*	.75	.75	2.57
Session x message	1	.08	.08	1.10	.33	1.09		.26	.26	1.26	1.02	1.02	3.49
Method x message	1	1.51	1.51	19.66*	3.00	3.00	9.81*	3.52	3.52	17.38*	1.88	1.88	6.43*
Error	20	1.53	.08		6.11	.31		4.05	.20		5.84	.29	
Within													
Session	1	.02	.02		.26	.26	1.18	.42	.42	1.25	.19	.19	.42
Method	1	.13	.13		32.51	32.51	150.42*	17.52	17.52	51.99*	7.92	7.92	17.87*
Message	1	.63	.63		.08	.08		.19	.19	.56	.01	.01	.01
S x M x M	1	.08	.08		1.34	1.34	6.18*	1.50	1.50	4.47*	.52	.52	1.18
Error	20	3.07	.15		4.32	.22		6.74	.34		8.86	.44	

		TRIAL								
		Final A vs 4B			Final A vs 5B			Final A vs 6B		
Source of Variance	df	SS	MS	F	SS	MS	F	SS	MS	F
Between										
Groups	3	4.18	1.39	4.11*	4.18	1.39	3.75*	3.99	1.33	2.70
Session x method	1	.57	.57	1.69	.26	.26	.69	.42	.42	.85
Session x message	1	1.42	1.42	4.18	1.17	1.17	3.15*	1.69	1.69	3.42
Method x message	1	2.19	2.19	6.46*	2.76	2.76	7.41*	1.88	1.88	3.81
Error	20	6.79	.34		7.44	.37		9.86	.49	
Within										
Session	1	.11	.11		.01	.01	.01	.05	.05	.23
Method	1	4.85	4.85	12.65*	2.76	2.76	8.14*	1.02	1.02	4.89*
Message	1	.001	.001		.05	.05	.14	.01	.01	.03
S x M x M	1	.30	.30		.42	.42	1.25	.19	.19	.89
Error	20	7.66	.38		6.77	.34		4.18	.24	

*p < .05

Comparison of performance on the first trial of each work method revealed no significant differences, indicating that methods A and B had comparable baselines. The percentage of items correctly reported on the final trial of method A and the fourth and fifth repetitions of method B did not differ significantly. However, the final trial of method A received a significantly higher quality rating than all trials of method B ($p \leq .05$).

Groups of subjects were found to differ significantly in all comparisons except the quality rating for the final trial of method A and the fifth repetition of method B ($p \leq .05$). In the experimental design employed, these group differences are confounded with two-way interactions. Most of the variation between groups corresponds to the interaction between methods and messages. Study of such interactions was not a major goal of the present experiment, and only limited attention is given to the present result.

The method x message interaction is shown in Figure 2. With method A, performance was superior on message 1, while with method B performance was superior on message 2. This result suggests that the best work method may be determined by characteristics of the message. It is not clear what these determining characteristics are. As may be seen in Figure 2, there was very little difference in performance with methods A and B on message 2.

The percentage of items correctly reported in message 1 was significantly smaller than in message 2 when the first and final trials of method A were compared to the first trial of method B ($p \leq .05$). Message differences were not significant in any other comparisons.

The significant interactions observed in Tables 2 and 3, with the exception of method x message interaction, were not consistent across all analyses and are not important to this discussion.

Method A produced a superior product in a shorter time than did method B. A complete gist was prepared in 28 minutes average time using method A. With method B, where the time was controlled, a gist was turned in approximately every 8 minutes following each repetition. Thus, it required approximately 48 minutes to produce the final gist. Comparisons of methods A and B for percent items correctly reported indicated that there was no significant difference between the fourth trial of method B and the final trial of method A. The difference in time to arrive at this point was four minutes, in favor of method A.

Repetition

As expected, performance scores improved as a direct function of repetition. A trial x group of subjects analysis of variance was computed for methods A and B for each performance measure. A summary of the analyses of variance is provided in Table 4. Performance with method A increased significantly from a mean of 42.5 percent words correctly

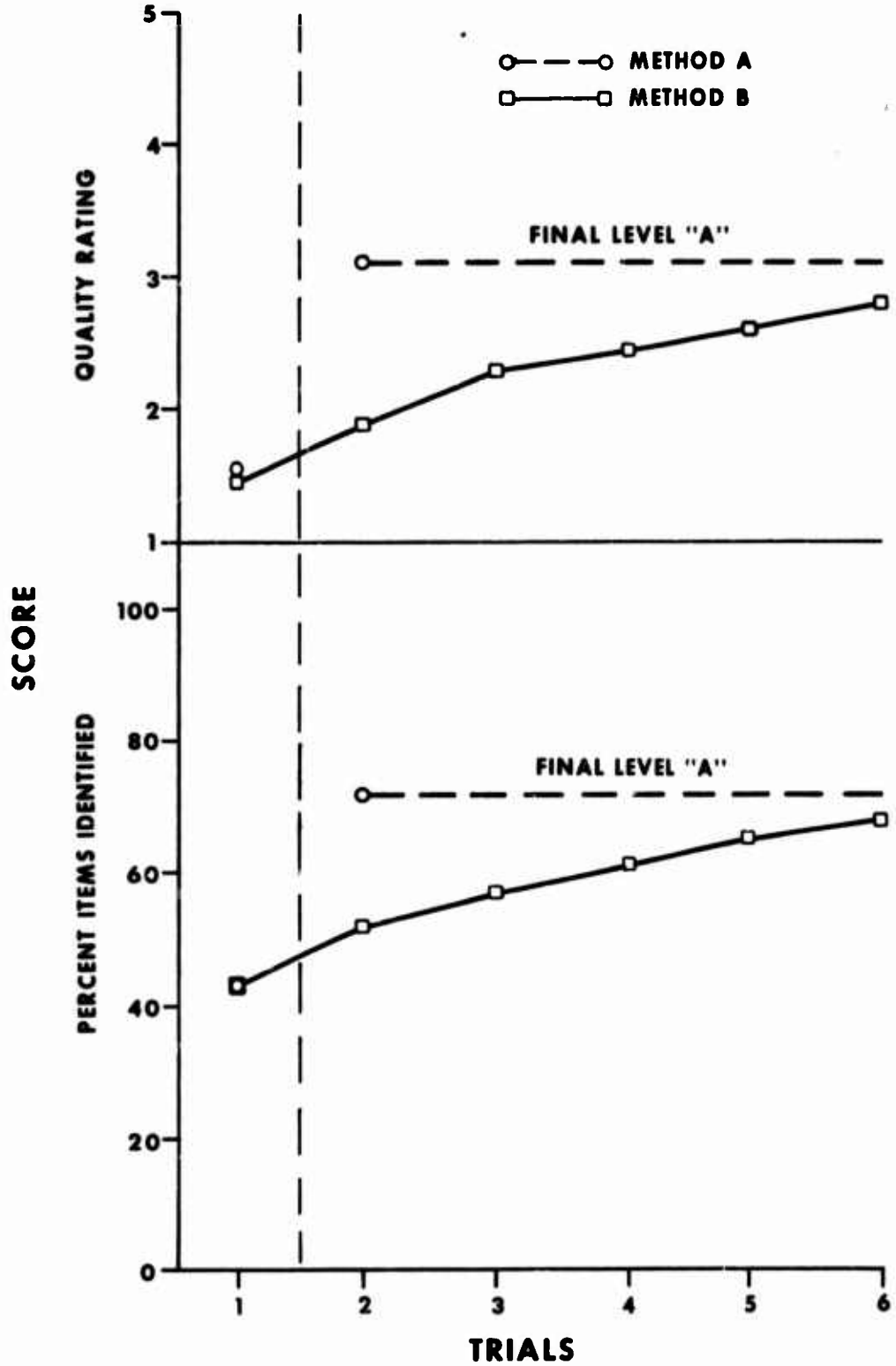


Figure 1. Comparison of average gisting scores for the first and final trial of method A with the six trials of method B for percent items correctly identified and quality rating.

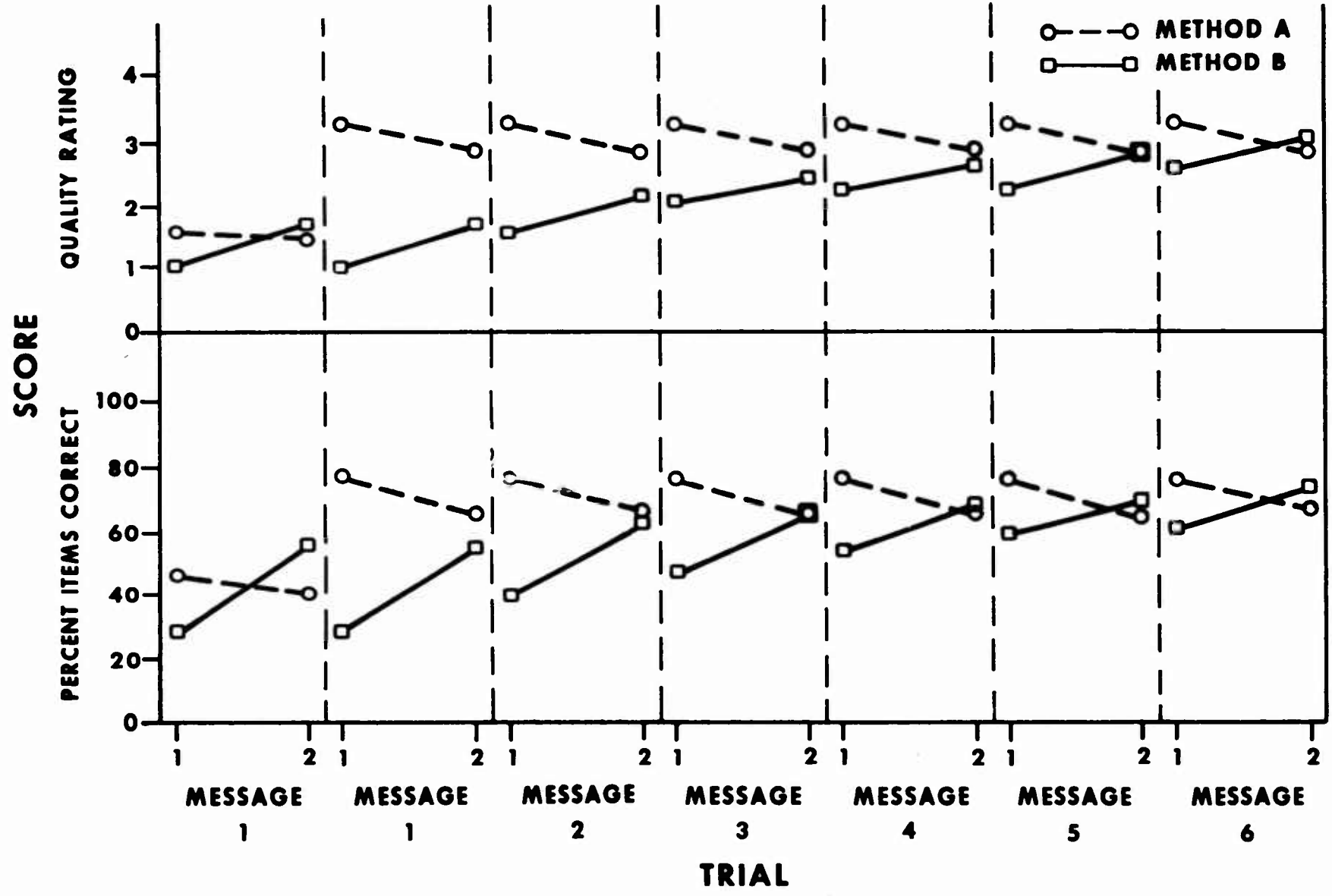


Figure 2. Average gisting scores for messages 1 and 2 by trials for methods A and B for percent items correctly identified and quality rating. The final trial of method A is compared to each trial of method B.

identified to a mean of 72.0 percent in the final gist, and from a mean quality rating of 1.6 on the first gist to 3.1 on the final gist.

With method B, the mean percentage of words correctly identified increased from 42.6 percent on the first gist to 67.8 percent on the sixth gist, and the mean quality rating increased from 1.5 to 2.8. Since the analysis of variance revealed a significant main effect for repetition (Table 2), individual mean differences between repetitions of method B were assessed by the Newman Keuls technique for each performance measure[2]. The results of the analyses are shown in Table 5. There were no significant differences in the percentage of items correctly identified between trials 2 and 3, 3 and 4, 4 and 5, and 5 and 6. The mean increase in quality rating was not significant between repetitions 3 and 4, 4 and 5, and 5 and 6. All other differences were significant.

Using scores on the first gist with each method as a base-line measure of performance, percentage gain in percent words correctly identified and percentage gain in quality ratings as a function of repetition was computed using the previous repetition as a base-line. The percentage gains are shown in Figures 3 and 4. In addition, percentage gain with method A across trials using the first trial as a base-line have been computed separately for messages 1 and 2 and are shown in Figure 5.

The number of tape reversals were counted for 15 of the 24 gisters. The frequency of reversals varied from 5 to 258. Fourteen of the 15 gisters made between 5 and 33 reversals. These scores were used to compute correlation coefficients between number of reversals and 1) percent items correctly identified, 2) quality rating and 3) time to complete gist. The correlation coefficients were .31, .39 and .44, respectively. It does not appear that frequency of reversal is a good predictor of performance.

DISCUSSION

The two work methods of gisting examined in this study were designed so that the effect of repetition on the accuracy and quality of the gist could be determined. As expected, accuracy increased as a direct function of repetition (Figure 1). However, it appears that the complete repetition of a five-minute message with no opportunity to stop or repeat smaller segments (method B) is not an economical method of gisting. With method A, where the gister was allowed to repeat segments of his choice, a final gist of higher quality was prepared in a shorter time.

With method A, an average of 72 percent of the items were identified correctly in the final completed gist. The final product was produced in approximately 28 minutes. The gist completed after one complete presentation of the message -- the mode which simulates on-line gisting-- was prepared in approximately 8 minutes, and an average of 42 percent of the

---

[2] See footnote 2, page 3.

Table 4

SUMMARY OF ANALYSIS OF VARIANCE OF TRIALS BY GROUPS FOR EACH PERFORMANCE MEASURE FOR METHODS A AND B

METHOD A

		Percent Items Identified			Quality Rating		
Source	df	SS	MS	F	SS	MS	F
Between	23	7543.7171			8.0417		
Groups	3	1791.5527	597.1842	2.076	2.2604	.7535	2.607
S within G	20	5752.1644	287.6082		5.7813	.2891	
Within	24	13621.9291			34.7500		
Trials	1	10463.3649	10463.3649	82.849*	28.5208	28.5208	111.984*
G x T	3	632.6614	210.8871	1.670	1.1354	.3785	1.486
e_w	20	2525.9028	126.2951		5.0938	.2547	

METHOD B

		Percent Items Identified			Quality Rating		
Source	df	SS	MS	F	SS	MS	F
Between	23	24233.9955			26.7287		
Groups	3	14054.5973	4684.8658	9.205*	13.0326	4.3442	6.344*
S within G	20	10179.3982	508.9699		13.6962	.6848	
Within	120	19543.6139			46.9479		
Trials	5	10483.7084	2096.7417	29.144*	30.2000	6.0400	40.772*
G x T	15	1865.5798	124.3720	1.729	1.9336	.1289	.870
e_w	100	7194.3258	71.9433		14.8142	.1481	

*$p \leq .05$

Table 5

MEAN DIFFERENCES BETWEEN TRIALS OF METHOD B FOR PERCENT ITEMS CORRECTLY IDENTIFIED AND QUALITY RATINGS USING THE NEWMAN-KEULS PROCEDURE

		Percent Items Correctly Identified						Quality Rating					
Means		1 42.62	2 51.92	3 56.80	4 61.47	5 65.36	6 67.81	1 1.458	2 1.896	3 2.292	4 2.469	5 2.625	6 2.813
Order		6	5	4	3	2	1	6	5	4	3	2	1
Trial	1	25.19*	22.74*	18.85*	14.18*	9.30*	-	1.355*	1.167*	1.011*	0.834*	0.438*	-
	2	15.89*	13.44*	9.55*	4.88	-		0.917*	0.729*	0.573*	0.396*	-	
	3	11.01*	8.56*	4.67	-			0.521*	0.333*	0.177	-		
	4	6.34*	3.89	-				0.344*	0.156	-			
	5	2.45	-					0.188*	-				
	6	-						-					

*Significant at the .01 level.

items were identified correctly. That is, it required three and a half times longer to produce an absolute gain of 30 percent -- or, using the first trial as the baseline, a relative gain of 69 percent items correctly identified. On the first presentation of the message, the average quality rating was 1.5 (midway between unacceptable and poor), and on the final version, the quality rating increased to 3.1 (fair), or a percentage gain of 99 percent. The acceptable trade-off between timeliness and accuracy probably varies with the specific situation. Therefore, in some situations, the on-line gist may satisfy the requirement for timeliness.

With method B, there were two constraints: 1) The complete message was repeated with no opportunity to stop and repeat small segments; and 2) only three minutes were allowed after each trial to complete the gist. The time constraint between trials may have been a limiting factor. It may be that gists of higher quality would have been prepared if more time had been given between trials. As shown in Figure 3, performance continued to improve through the fifth repetition. However, percent gain in words correctly identified leveled off after the second repetition (Figure 4).

With method B, it is possible to estimate the gain in information per unit time. Each trial required approximately 8 minutes to complete. Therefore, the time represented by the second trial is twice that of the first, or 2 to 1, and the third is 3 to 1, and so on for each trial, using the first trial as the baseline. As shown in Figure 4, there was a 22 percent gain in the percentage of items correctly identified and a 30 percent gain in quality rating on the second trial or in twice the time as the first trial. Two complete presentations may be a feasible procedure for producing a quick gist.

The percent gain in performance varied with messages (Figure 5). Obviously, some messages profit more from repeated listening than others. On the basis of data collected in the present study, it is difficult to account for these differences between messages. Part of the difference may be explained by differences in groups of subjects; however, it is doubtful that this difference accounts for the entire effect. The difference in messages makes it difficult to generalize about the effect of repetition. In some cases, repeated listening produced large gains, while in others, the percent gain in performance was relatively small.

The difference between methods A and B varied with the specific message (Figure 2). For one message, the highest scores were obtained with method A, and for the other message, the highest scores were obtained with method B. It was not the purpose of the present study to specify the properties of messages which determine differences in performance; therefore no explanation of this interaction is provided in these analyses. Messages appear to vary along many dimensions, and subtle differences between messages in all probability account for these observed interactions. The data collected cannot explain why one method is better for one message and another method is better for another. The

messages used in the present study were chosen to represent typical messages processed by communications operators. The two messages were equal in intelligibility, equally long, and relatively free from background noise. However, they varied along a number of other dimensions. The topics of conversation were different, the number of speakers varied, and the rate of speaking may have been unequal. Other more subtle differences more difficult to measure may also have existed, such as number of inferences contained in one message or directness of expression. The results of the present study underline the importance of identifying the properties of a communication which determine the difficulty of gisting and make one message different from another. The importance of these difficulties in identifying the properties of messages which determine performance has been substantiated by other studies (Quantitative evaluation of current procedures in voice processing, Technical Research Report 1174[4] and a study in progress on comparative gisting ability of LeFox Grey and standard systems personnel).

With regard to the general level of performance, the average quality rating obtained on the final gist with method A was 3.1, a rating of fair. This was the highest average obtained. An average of 72 percent of the items were identified correctly on the final completed gist with method A--the highest average. These scores were obtained on clearly spoken English language messages, and all speakers were native English language speakers. It is clear from these results that gisting involves more than language speaking ability.

In summary, the results imply that 1) the best method of gisting varies with the type of material; 2) there is need to identify the properties of a message which determine its difficulty; and 3) on-line gisting is a feasible consideration for timely reporting of the contents of a message.

---

[4] House, Joyce L. and Stanley L. Cohen. Quantitative evaluation of current procedures in voice processing (U). Technical Research Report 1174 (AD 516989). Behavior and Systems Research Laboratory. July 1971. CONFIDENTIAL

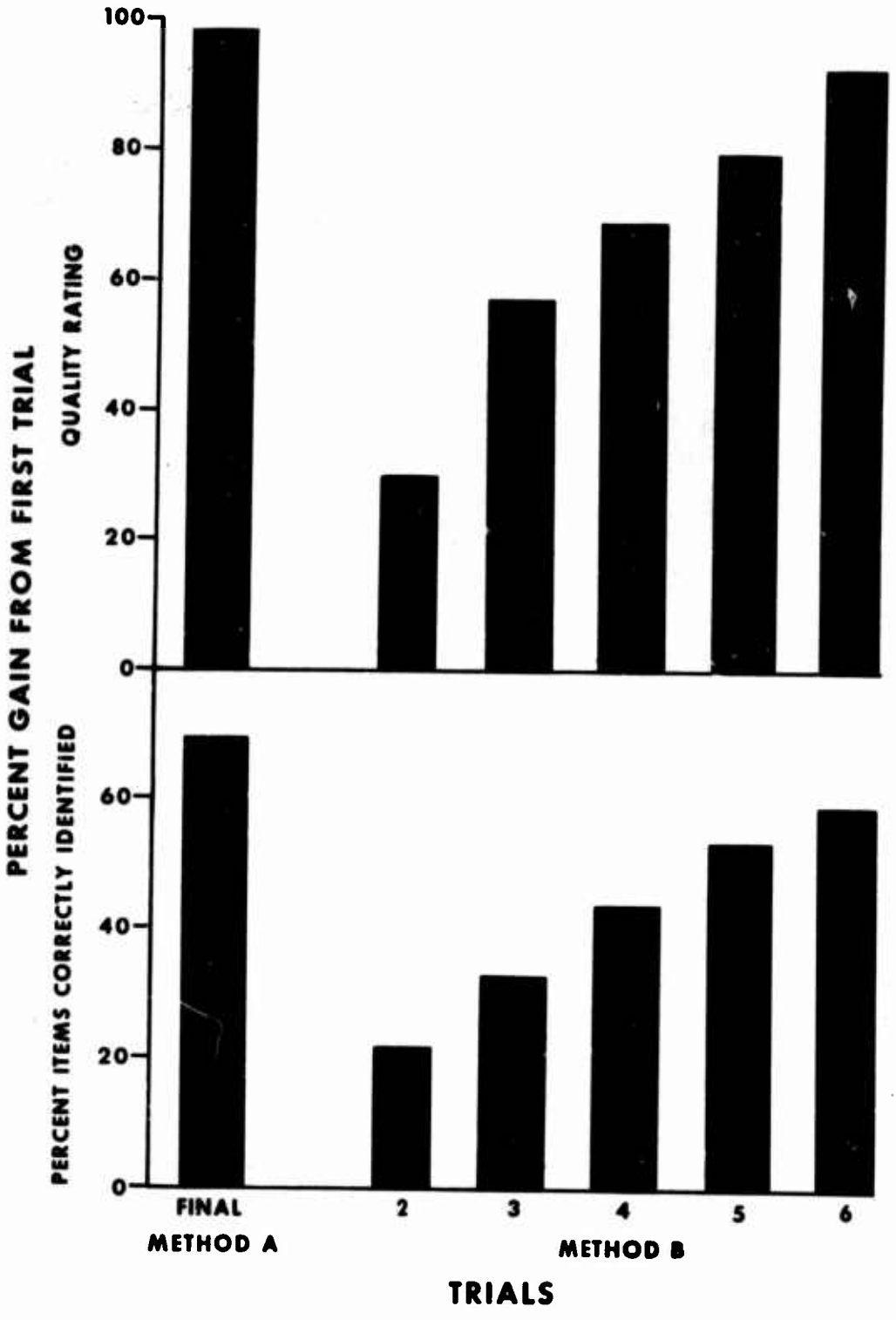


Figure 3. Average percentage gain in gisting score by trial by methods A and B for percent items correctly identified and quality rating.

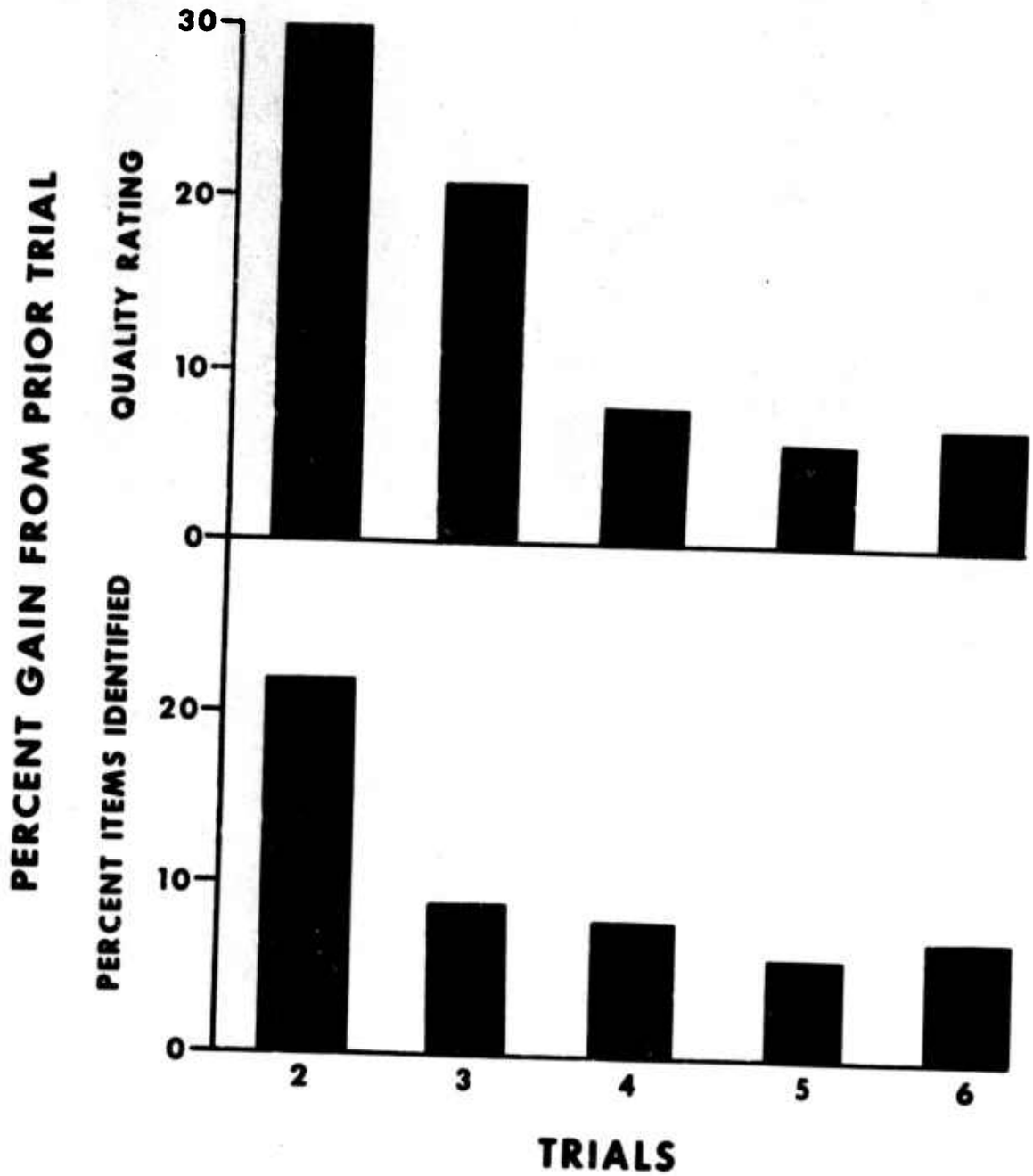


Figure 4. Average percentage gain in gisting score from previous trial by trial for method B for percent items correctly identified and quality rating.

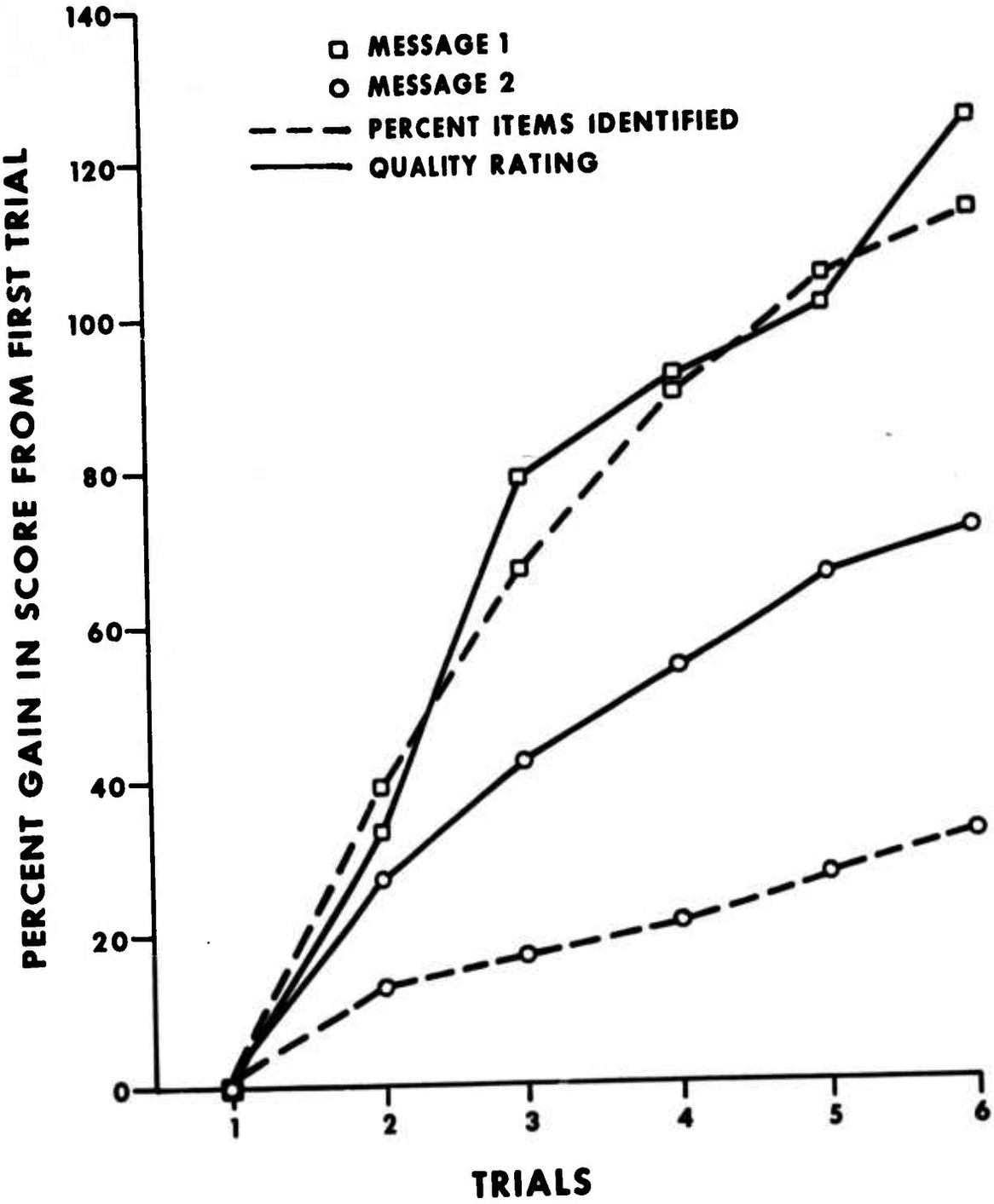


Figure 5. Average percentage gain in gisting score by trial for method B for percent items correctly identified and quality rating.

APPENDIX

FREQUENCY DISTRIBUTION OF SCORES FOR METHODS A AND B FOR PERCENT ITEMS CORRECTLY IDENTIFIED AND QUALITY RATING (N = 24)

	Percent Items Correctly Identified									Quality Rating								
	Trials									Trials								
	Method A		Method B							Method A		Method B						
Score	1	final	1	2	3	4	5	6	Score	1	final	1	2	3	4	5	6
95-100									5.00								
90-94		1							4.75								
85-89		4						2	4.50								
80-84		2				2	4	2	4.25		1						
75-79		4		1	2	3	2	2	4.00		2						
70-74		4		3	3	3	6	7	3.75		2					1	2
65-69	1	4	3	3	7	6	1	6	3.50		5			1	1	1	1
60-64	4		3	3	2	1	4		3.25		4			1	1	3	3
55-59	3	3	2	2	1	1	2	1	3.00		2		2	1	5	5	7
50-54			2	1	2	2	1		2.75		1	1	1	3	5	3	3
45-49	1	1	3	1	1	3	2	2	2.50		3			6	2	1	4
40-44	3			4	2	1	2	2	2.25	2	2	1	4	3	3	3	1
35-39	5	1	1	1	2				2.00	4		3	6	4	3	5	1
30-34	4		3	3		2			1.75	3	2	3	3	2		2	1
25-29	1		2		1				1.50	7		3	4	1	4		1
20-24			2	1					1.25	5		5	2	1			
15-19			2	1					1.00	3		8	2	1			
10-14	1				1												
5-9			1														
0-4	1																